

VU Research Portal

Reflexity in teams: A measure and correlates.

Schippers, M.C.; den Hartog, D.N.; Koopman, P.L.

published in

Applied Psychology
2007

DOI (link to publisher)

[10.1111/j.1464-0597.2006.00250.x](https://doi.org/10.1111/j.1464-0597.2006.00250.x)

document version

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

citation for published version (APA)

Schippers, M. C., den Hartog, D. N., & Koopman, P. L. (2007). Reflexity in teams: A measure and correlates. *Applied Psychology*, 56(2), 189-211. <https://doi.org/10.1111/j.1464-0597.2006.00250.x>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl

Reflexivity in Teams: A Measure and Correlates

Michaéla C. Schippers* and Deanne N. Den Hartog

Erasmus University Rotterdam, The Netherlands

Paul L. Koopman

Free University Amsterdam, The Netherlands

Reflexivity—the extent to which teams reflect upon and modify their functioning—has been identified as a possible key factor in the effectiveness of work teams. The aim of the present study was to develop a questionnaire to measure (aspects of) reflexivity, with a focus on team reflection. The questionnaire was tested in two different samples, namely a first sample of 59 teams from 14 different organisations (Study 1) and a confirmation sample of 59 school management teams (Study 2). In both samples, two factors of reflection were identified. These were labeled evaluation/learning and discussing processes. Scale statistics showed good psychometric properties for the scales in both studies. We conclude that the scales form a parsimonious and valid instrument to assess reflexivity in teams.

La réflexivité (dans quelle mesure les équipes remettent-elles en cause leur fonctionnement et le modifient-elles?) est apparue comme étant l'un des éventuels facteurs clés de l'efficacité des groupes de travail. Cette recherche s'est donné pour tâche de développer un questionnaire pour évaluer certaines dimensions de la réflexivité en mettant l'accent sur la réflexion de groupe. Le questionnaire a été mis à l'épreuve auprès de deux échantillons, à savoir un premier échantillon composé de 59 équipes relevant de quatorze organisations différentes (Etude n°1) et un second échantillon de confirmation réunissant 59 équipes d'écoles de management (Etude n°2). Deux facteurs relatifs à la réflexion sont apparus dans chacune des deux études. Ils ont été dénommés évaluation/apprentissage et processus de discussion/principes. Dans les deux échantillons, les échelles ont obtenu des résultats statistiques satisfaisants sur le plan psychométrique. On arrive à la conclusion que ces échelles constituent un outil valide et économique pour évaluer la réflexivité dans les équipes.

* Address for correspondence: Michaéla C. Schippers, Erasmus University Rotterdam, Rotterdam School of Management, Burg. Oudlaan 50, 3062 PA Rotterdam, The Netherlands. Email: mschippers@rsm.nl.

Deanne Den Hartog is now at the University of Amsterdam, The Netherlands.

The help of consultancy firm Beteor in obtaining the sample for the second study was much appreciated. The authors also wish to thank Daan van Knippenberg and three anonymous reviewers for their helpful comments on an earlier version of this manuscript.

INTRODUCTION

Effective teams are important cornerstones of successful organisations, especially for those operating in dynamic environments. West identified reflexivity as an important determinant of team effectiveness (West, 1996). When members collectively reflect on the way they work and the environment they work in, plan to adapt these aspects and make changes accordingly, teams will be more effective (West, 2000).

Although reflexivity can be operationalised at the individual level (e.g. as a cognitive style, cf. Petzold, 1985), it is more typically thought of as a group process. Team reflexivity is defined as “the extent to which group members overtly reflect upon, and communicate about the group’s objectives, strategies (e.g. decision-making) and processes (e.g. communication), and adapt them to current or anticipated circumstances” (West, Garrod, & Carletta, 1997, p. 296). Non-reflexive teams show little awareness of the team objectives, strategies, and the environment in which they operate. Such teams are inclined to be reactive rather than proactive and react defensively in case of environmental threat. Reflexive teams show more detailed planning, pay more attention to long-term consequences, and have a larger inventory of environmental cues to which they respond (West et al., 1997).

To date, relatively little research on reflexivity has been done. Carter and West (1998) found that reflexivity predicted the effectiveness of BBC production teams. Schippers, Den Hartog, Koopman, and Wienk (2003) found that group longevity and outcome interdependence moderated the relationship between diversity on the one hand and reflexivity and team outcomes on the other. De Dreu (2002) found that high levels of minority dissent led to more innovations and greater team effectiveness, but only if the level of team reflexivity was high.

As stated, research on reflexivity is limited and there are also only few measures available. Although reflexivity is thought of as an iterative process consisting of several components (reflection, planning, and action/adaptation, e.g. West, 1996, 2000) and the reflection component is assumed to have three levels of depth (shallow, moderate, and deep; Swift & West, 1998), most studies measure reflexivity as a unidimensional construct, using a short questionnaire developed by Swift and West (1998). This measure does not tap different levels of reflection and does not include items on adaptation. The main aim of this study is thus to develop a more elaborate questionnaire to measure reflexivity, and especially reflection in teams. Although it is intuitively appealing that teams should reflect in order to work more effectively, this has received little attention in previous research. Thus, in developing the questionnaire, we focus on the reflection component. Here, we take reflection as the starting point of the reflexivity process, and we are only interested in the action/adaptation component insofar as it is linked to

or follows from reflection. A second aim of the current research was to explore the relationships between reflexivity and several related constructs, such as feedback-seeking behavior and the reflector learning style.

Two studies were done using separate samples of teams. The second study is used to cross-validate the new instrument developed in the first study, and analyses presented here include assessments of discriminant validity, reliability, and factor structure. Before elaborating on the studies, the components of reflexivity are described in more detail.

Reflection

Team reflection refers to a team's joint and overt exploration of work-related issues. According to West (2000, p. 4), "reflection includes behaviors such as questioning, planning, exploratory learning, analysis, diversive exploration, making use of knowledge explicitly, planfulness, learning at a meta-level, reviewing past events with self-awareness, and coming to terms over time with a new awareness". Reflection helps recognise how present ways of operating may have become obsolete due to environmental changes (Tjosvold, 1991).

Reflection can occur before, during, or after execution of the team task and can vary in time scale from short term to long term (West, 2000). Reflection before task execution is characterised by joint consideration of team goals, strategies, and processes. This includes reflection on the nature of the problem that faces the team (Moreland & Levine, 1992). Reflection during team task execution implies reviewing whether the team is still on track, whether the right problem is being solved, and whether things are done correctly. Reflection after finishing the task implies evaluation of achievements and the way things were done.

Reflection can also vary in depth (Swift & West, 1998). *Shallow reflection* is the first phase of awareness and involves thinking about issues related closely to the task at hand, for example, discussing the division of tasks among team members. Shallow reflection is similar to so-called single-loop learning (Argyris, 1992). *Moderate reflection* is characterised by a more critical approach toward tasks, goals, strategies, and processes. It is similar to double-loop learning (Argyris, 1992). In the phase of *deep reflection*, the norms and values of the team or organisation are questioned and their effect on team and organisational functioning is discussed. This phase is similar to what is called generative (Senge, 1990) or triple-loop learning (Nielson, 1993; Snell & Chak, 1998). Deep reflection may be more important for specific types of teams. For instance, reflecting on the culture of the organisation as a whole and its impact may be more relevant for management than for production teams. Deep reflection is expected to take place less often than shallow and moderate reflection as most teams will tend to take culture for granted and will not discuss cultural norms and values on a regular basis (Allen, 1996).

Adaptation

Reflection as such does not lead to changes. Some adaptation needs to occur as well. According to West (2000, p. 6), the action/adaptation stage refers to “goal-directed behaviors relevant to achieving the desired changes in team objectives, strategies, processes, organisations or environments identified by the team during the stage of reflection”. West (2000) asserted that action or adaptation could be measured on four dimensions: magnitude, novelty, radicalness, and effectiveness. The first three dimensions describe the innovativeness of the action, whereas the fourth is related to the performance of the team. Adaptations carried out by the team lead to new information, which can lead to further reflection, planning, and adaptation as an iterative and ongoing process (West, 2000). However, not all teams have innovative goals, and when having a more diverse sample in which teams will have more diverse goals it may be more relevant to see adaptation as the extent to which intended changes, detected or planned during the reflection phase, are carried out. We therefore chose to define and operationalise adaptation as the extent to which teams live up to agreements.

Related Constructs

Constructs we expect to be related to team reflexivity include feedback-seeking behavior, the level of proactive personalities within the team, as well as the level of the reflector learning style within the team.

Feedback-seeking Behavior. Frese and Zapf (1994) define feedback as “information about how far one has progressed toward the goal” (p. 279). Usually this goal is related to job performance. Employees seek feedback through active monitoring and inquiry of information to facilitate achievement of job performance and other goals important to them (Ashford & Cummings, 1983; for a recent review, see Ashford, Blatt, & VandeWalle, 2003). Feedback-seeking behavior is closely related to reflexivity in that engaging in behavior that helps to obtain information to reflect on can be important to enable (accurate and relevant) reflection. The amount of feedback seeking and the way the obtained information is handled by the team can be seen as an indicator of team reflexivity (West et al., 1997). However, although feedback-seeking behavior is important for reflexivity, it is not identical to reflexivity. Reflexivity has to do with how things can be improved, while feedback seeking is getting information on how far one is from the (performance) goal and does not necessarily imply that the obtained information is reflected upon. Ancona and Caldwell (1992) showed that actively seeking information and feedback outside the team is related to team success. In the current study then, we expect positive relationships

with the different elements of reflexivity (i.e. reflection and adaptation), especially reflection.

Proactive Personality. According to Bateman and Crant (1993, p. 105), a person with a proactive personality is one who “is relatively unconstrained by situational forces, and who effects environmental change”. People low on this trait can be characterised as more passive, they tend to “react to, adapt to and be shaped by their environments”. In order to effect environmental change, challenging existing norms and values is needed. We thus expect teams with relatively more proactive individuals to show more reflection as there are more individuals in the team who are likely to “speak out”. However, being reflexive as a team also depends on other variables, such as the sense of psychological safety (Edmondson, 1999) and trust. Thus, although we expect positive correlations between reflexivity components and proactive personality (in terms of the relative level of this trait within teams), we expect these correlations to be fairly low, as team reflexivity should differ from the level of proactive personalities within a team.

Learning Styles. Another personality trait that could be important for the way people deal with information, and thus reflexivity, are individual learning styles. According to Kolb (1984), people differ in the way they learn from experience. Kolb (1984) stated that the learning cycle has four phases, namely, experiencing, reflecting on the experience, theorising, and active experimentation. In line with this theory, Honey and Mumford (1995) distinguished between four learning styles, namely reflector, theorist, pragmatist, and activist. Of these, the reflector learning style (reflecting on experience) seems especially relevant in the context of reflexivity. Although teams with more individuals high on the reflector learning style will have a higher potential to be reflexive in that there will be more people prone to individual reflection, actual levels of team reflexivity will also depend on other variables, including the extent to which people are willing and able to voice their thoughts in the team. We therefore expect positive correlations between reflexivity components and the reflector learning style (in terms of the relative level of this trait within teams). However, we expect these correlations to be low, as team reflexivity as a team-level process should differ from the mean level of this individual learning style.

METHOD

Sample and Procedure

Fifty-nine teams (454 persons) from 14 different organisations participated in this study. The teams included management teams, self-regulating teams,

production teams, teams in government service, executive teams in schools, and facilitating teams. The teams came from companies in the IT, insurance/banking sector, government, and chemical industry. Drawing on Hackman (1987) we considered teams to be composed of individuals who both see themselves and are seen by others as an interdependent social entity. Teams are embedded in a larger organisation, and their performance affects others, for instance suppliers or customers. Only teams that met these criteria were included. We checked this when first speaking to a contact person and again when meeting the teams. In most teams, members were assigned to the teams when they were first formed, thus teams did not select members themselves. Tasks differed widely, from administrative or production work (production teams) to leading a company (management teams). Only teams with highly routine jobs were not included, as reflection seems less relevant for such teams.

The initial sample of teams that agreed to participate consisted of 60 teams. The response rate within the teams that had agreed to participate was 91 per cent. Two questionnaires were incomplete and thus excluded from further analyses, and one team was excluded because only one team member returned the questionnaire. The remaining respondents ($N = 453$) were from 59 teams ranging in size from four to 22 members with an average of 7.68 persons per team and at least two respondents per team. In most teams, all team members returned the questionnaire, and at least 50 per cent of the team members returned the questionnaire. The mean age of respondents was 39.5 years ($SD = 9.37$) and 64 per cent were male.

Teams were recruited by phone. For 33 teams, questionnaire packages were mailed to the team leaders who had agreed to participate. Leaders then handed questionnaires to team members. A cover letter described the purpose of the study and guaranteed confidentiality. Completed questionnaires were sent directly to the researchers. In 26 teams, a researcher went to the workplace and handed out questionnaires, respondents filled it out and handed it back.

Measures

Items were written to tap all three levels of reflection, namely shallow, moderate, and deep. As stated, most research on team reflexivity has used a nine-item unidimensional measure developed by Swift and West (1998). However, this measure does not tap levels of depth and does not contain items of adaptation, which are also part of the construct of reflexivity. We decided to develop an extended measure to tap the hypothesised subscales of reflection related to levels of depth. Also, we saw evaluating and learning from actions and adaptations as an important component of reflection and items on this were developed and included. Items formulated by Swift and West (1998) were also translated and included in the pool of 34 items measuring the reflection component. Furthermore, items were formulated to cover the domain of

adaptation (five items). The items for adaptation were formulated after interviewing two team leaders and asking them what they thought was important in the context of adaptation. Both managers pointed out that it was important that actions/adaptations agreed upon were actually carried out. One of the managers complained that often what was agreed upon was not implemented, even if the agreements were written down. Finally, questions were formulated to measure feedback-seeking behavior (five items).

Based on comments of three experts in the field of organisational psychology, several items were rephrased or left out. Next, a short pilot study was done among 28 members of two teams of HR managers from a large IT company. The questionnaire was given to the first team and all 16 members filled out the questionnaire and provided comments. Most items were understood, and seen as unambiguous and relevant. Some adaptations were made based on the comments before giving the questionnaire to the second team. All members filled it out and commented on the items, resulting in a few additional minor changes.

The Final Questionnaire

Reflection. The final reflection scale consisted of 34 items. Nine of these (items 26 through 34) were from the reflexivity scale of Swift and West (1998). The first level of depth is shallow reflection, which involves teams thinking about and discussing issues related closely to the task at hand. An example of an item intended to measure this level is: "Before we start working on a task, we take time to decide on the best working method". The second level of depth is moderate reflection, where teams take a more critical approach toward their work processes. An example of an item is: "We regularly discuss whether the team is working effectively together". The third level of depth, deep reflection, entails questioning the prevalent norms and values. An example of an item intended to tap this level is: "This team is prepared to challenge organisational practices and policies" (see the Appendix for the full set of items). Responses are given on a 5-point scale (1 = strongly disagree, 5 = strongly agree).

Adaptation. Five items tap the extent to which the team members carry out planned actions and make adaptations that were agreed upon, for example: "In this team agreed upon actions are usually carried out" (1 = never, 5 = always), $\alpha = .79$. The Appendix lists all items.

Feedback-seeking Behavior. Five items assess the extent to which team members actively seek feedback on their method of working, for example: "We seek feedback on our methods" (1 = never, 5 = always), $\alpha = .73$. The Appendix lists all items.

Proactive Personality and Learning Style Reflector. Proactive personality was measured using eight items of the proactive personality scale developed by Bateman and Crant (1993). An example of an item is: "I am always looking for better ways to do things" (1 = strongly disagree, 5 = strongly agree), $\alpha = .73$. The learning style reflector was measured using three items from the scale developed by Honey and Mumford (1995). The original items were Q-sort, so to fit better with the other items used in our study we reformulated them using Likert-type scales. The items were: "I like to reach a decision after carefully weighing up many alternatives", "I am careful not to jump to conclusions too quickly", and "I prefer to have as many sources of information as possible—the more data to think over the better" (1 = strongly disagree and 5 = strongly agree), $\alpha = .68$.

RESULTS

Factor Analysis

An exploratory factor analysis was done, using an oblique factor solution. Since we expected the extracted factors to be correlated rather than independent, the use of an oblique factor solution (oblimin rotation) was more adequate than an orthogonal one. In the initial solution, seven factors with an eigenvalue above 1 were found. Using the "scree criterion" (Cattell, 1966), two interpretable factors explaining 39 per cent of the variance emerged. The factors resembled shallow and moderate reflection. They were named *evaluation/learning* and *discussing processes*, as these labels better reflected item content (see Table 1).

To assess reliability and internal consistency, several criteria were used. Items that did not meet these criteria were left out in subsequent analyses to increase the homogeneity of the scales. The first criterion used is that the Cronbach's alpha should be $> .70$ (Nunnally, 1976). Factor loadings should be $> .40$, the difference between factor loadings of an item should be $> .20$ and item-rest correlations should be $> .30$ (e.g. Den Hartog, Van Muijen, & Koopman, 1997). Ten items did not meet criteria and were discarded. One deep reflection item loaded on factor one, instead of factor two (item 34), and was discarded for theoretical reasons. The final scales consisted of 19 and four items respectively. The deep level items loaded mostly on the second factor, and some were discarded, because they had high factor loadings on both factors.¹ All items of the second subscale (labeled discussing processes) are from the Swift and West (1998) scale. Table 1 reports the factor loadings of the items. Table 2 presents *F*-values and ICC values as well as Cronbach's α and team-level intercorrelations between all scales.

¹ The a priori items are available on request from the first author.

TABLE 1
Loadings on the Items¹ of Reflection on the Factors Evaluation/Learning and
Discussing Processes Respectively (Study 1)

	<i>Factor 1</i>	<i>Factor 2</i>
1. Take well-considered decisions	.61	.04
4. Review method of working	.54	.13
5. Talk about different ways of reaching objective	.72	.21
6. Only talk about critical/serious problems (R)	.44	-.09
7. <i>Discuss routines</i>	.60	.40
8. Examine implications of changes	.60	.23
9. Learn from past activities	.62	.18
10. Same problem definition	.63	.20
11. Stop to assess whether team is on right track	.64	.23
12. Talk about problems with team members	.59	-.02
13. Examine long-term consequences	.67	.12
14. Question objectives	.65	.32
15. Consider different points of view	.61	.13
17. <i>Examine contribution to organisation</i>	.56	.39
18. Evaluate whether activities produced expected result	.66	.32
19. Evaluate results	.68	.23
20. Reflect on developments	.58	.22
21. <i>Challenge norms and values team</i>	.53	.33
22. Evaluate things that don't work out as planned	.74	.12
23. Analyse success	.64	.39
24. Evaluate job done	.64	.30
25. Find cause of problems	.61	.25
29. <i>Adapt objectives to changing circumstances</i>	.59	.33
34. <i>Challenge organisational practices and policies</i>	.46	.09
2. <i>Reflect on way of communication</i>	.47	.50
3. <i>Reflect on way of decision making</i>	.58	.42
26. Reviewing objectives of team	.41	.67
27. Discuss methods used by team	.27	.72
28. Discuss if working effectively together	.46	.71
30. <i>Change of team strategies</i>	-.07	.20
31. <i>Discuss communication of information</i>	.43	.56
32. Reviewing approach of getting job done	.42	.66
33. <i>Alter way of decision making</i>	-.21	.11
Eigenvalue	10.9	1.9
Explained variance	33.1%	5.6%

Note: N = 454. PCA with oblimin rotation; Explained variance (cumulative): 39%; (R) = Recoded item.

¹ Short versions of the items are given. Full text items can be found in the Appendix. Items shown in italics are discarded. Item 16 was discarded beforehand, because it (theoretically) belonged to neither factor.

TABLE 2
Descriptive Statistics, *F*-values and ICC Values, and Team-Level Intercorrelations (Study 1; *N* = 59 teams; *n* = 453)

Variable		M	SD	<i>F</i> _(59, 392)	<i>ICC</i> (1)	<i>ICC</i> (2)	1	2	3	4	5	6
1.	Evaluation/learning	3.31	.38	3.62**	.27	.75	.91					
2.	Discussing processes	2.89	.39	2.39**	.18	.64	.53**	.77				
3.	Adaptation	3.35	.47	5.09**	.35	.80	.65**	-.07	.79			
4.	Feedback-seeking bh.	2.82	.40	2.62**	.18	.61	.68**	.51**	.66**	.73		
5.	Proactive personality	3.50	.24	1.33	.04	.17	.25	.36**	.07	.12	.73	
6.	Reflector	3.79	.31	1.45*	.07	.39	.22	.00	.24	.03	.17	.68

* *p* < .05; ** *p* < .01; two-tailed; Cronbach alphas are shown on the diagonal for all scales (in italic).

Levels of Analysis

Although two variables were measured at an individual level (proactive personality and the reflector learning style), the other variables are supposed to operate at the team level of analysis. In order to compute the correlations at the team level, the individual ratings are aggregated and reported at the team level. However, before aggregating individual-level scores, the viability of this procedure should be examined. Aggregating individual scores is only appropriate when between-group variance is significantly greater than the within-group variance and team members are sufficiently homogeneous in their scores (James, Demaree, & Wolf, 1984). One-way analysis of variance with the team as the independent variable and the scores of the team members for all the variables as the dependent variable (Kenny & La Voie, 1985) showed that, for all team-level variables, between-group variance was significantly greater than within-group variance. As expected, for proactive personality, and the reflector learning style, measured at the individual level, within-group variance was greater (see Table 2). In order to measure the extent to which teams differ with respect to proactive personality, the mean per team will be used to measure team personality elevation (Neuman, Wagner, & Christiansen, 1999).

Subsequently, ICC values (i.e. intraclass correlation coefficients assessing the relative consistency of responses among raters) were calculated (see Table 2). James (1982) reported a median ICC(1) (i.e. the degree of reliability associated with a single assessment of the group mean) of .12 for the organisational literature. For our team-level variables, ICC(1) ranged from .17 to .35, indicating that a considerable part of the variance is between groups. Especially for adaptation and evaluation/learning, an important part of the variance is between groups. For proactive personality and the reflector style, ICC(1) is well below .12. For ICC(2) (an estimate of the reliability of the group means) a criterion of between .60 and .70 is sometimes used for aggregation. However, ICC(2) depends also on team size, with higher values of ICC(2) as team size increases (Bliese, 2000). Therefore, although we do report ICC(2) values, we chose to depend mainly on ICC(1) in deciding on aggregation of individual-level scores.

Correlations and Discriminant Validity

As the pattern of correlations presented in Table 2 shows, the two reflection scales are significantly correlated. Most correlations are in line with expectations. The correlations also support the expectation that proactive personality is related to reflexivity; teams with more proactive individuals were higher on evaluation/learning and discussing processes. These correlations also show that reflexivity is related to but at the same time distinct from the

level of proactive personalities within the team. The learning style reflector was not significantly related to the subscales of reflection, supporting the idea that reflection on a team level differs from individual-level reflective learning styles. In order to get a first idea of the discriminant validity of the subscales of reflection, we tested whether the correlations of adaptation, feedback-seeking behavior, proactive personality, and reflector learning style differed for evaluation/learning and discussing processes, taking into account the intercorrelation between evaluation/learning and discussing processes (for a full description of this procedure, see Steiger, 1980). This was the case for all variables, with the exception of proactive personality, namely, $z = 6.31$, $p < .001$, for adaptation; $z = 1.77$, $p < .05$, for feedback-seeking behavior; $z = .90$, ns, for proactive personality; and $z = 1.72$, $p < .05$, for the learning style reflector. This provides a first indication of the discriminant validity of the subscales evaluation/learning and discussing processes. However, these analyses are explorative in nature. Therefore, we collected a cross-validation sample and conducted confirmatory factor analyses as well as several other tests to assess the discriminant validity of these newly developed scales.

STUDY 2

In the first study, a reflexivity measure was developed. Analyses resulted in two correlated yet interpretable subscales of reflection and a scale for adaptation, with satisfying internal consistency. Following the recommendations of Schwab (1980) and DeVellis (1991) to collect cross-validation data on a new sample when developing new scales, we conducted a second study for cross-validation purposes. This procedure has several advantages, e.g. assessing the dimensionality of the scale, avoiding sample-specificity, and avoiding unnecessary proliferation of constructs (Van Dyne, Graham, & Dienesch, 1994). The first study was conducted among teams with very diverse tasks. The second study uses a set of more homogeneous teams, namely school management teams (comprising school principals). We expected that the internal structure of the scale developed in Study 1 would also be obtained in Study 2.²

² As the proposed third dimension of reflection (i.e. questioning norms and values) was not found in the first study, seven items on this were added in an attempt to add a third scale to the operationalisation of reflection. However, the additional items all loaded on the second factor (discussing process) instead of the proposed third factor (discussing norms and values). We therefore chose to leave out these additional items and present Study 2 as an exact replication with respect to the items in Study 1. This also explains the somewhat irregular numbering of the items of reflection.

Sample and Procedure

The initial sample consisted of 60 teams (235 persons) from 51 high schools. The school management teams were found via the Internet, and were recruited by phone in collaboration with a consultancy agency. If teams agreed to participate, questionnaire packages were mailed to a contact person who handed the questionnaires to team members. A cover letter described the purpose of the study and assured confidentiality. Instructions for completion were given on the first page of the questionnaire. All individual team members returned the questionnaires directly to the researchers by mail. Two questionnaires were incompletely answered and excluded from further analyses, and one team was left out because only one team member returned the questionnaire. The remaining respondents ($N = 228$) were from 59 teams, ranging in size from three to eight members with an average of 3.7 persons per team and at least two respondents per team. Although in most teams all team members returned the questionnaire, overall at least 50 per cent of the team members returned the questionnaire. The overall response rate, within teams that had agreed to participate, was 87 per cent. Of these respondents, 81.6 per cent were male; 10 respondents did not state their gender. The mean age was 50.64 years ($SD = 6.07$); 27 people did not provide their age.

Measures

The items of the two subscales for reflection as developed in Study 1 and described above were included in the questionnaire. Furthermore, we included scales to measure adaptation, feedback-seeking behavior, and the reflector learning style (see Study 1 and the Appendix). Correlations and ICC-values are summarised in Table 3. For the team-level variables (evaluation/learning, discussing processes, adaptation, and feedback-seeking behavior), ICC(1) ranged from .14 to .24. Thus, a considerable part of the variance is between groups. For the reflector learning style, ICC(1) was .05, indicating that, as expected, most variance for this variable is at the individual level. The intercorrelations between the scales, presented in Table 3, are mostly similar to those found in Study 1.

Results

Confirmatory Factor Analysis. A confirmatory factor analysis was performed on the cross-validation data to assess whether the two previously found dimensions of reflection (evaluation/learning and discussing processes) would again describe the data well. We estimated the two-factor model using maximum likelihood techniques within LISREL VIII. The initial

TABLE 3
Descriptive Statistics, *F*-values and ICC Values, and Team-Level Intercorrelations (Study 2; *N* = 59 teams; *n* = 228)

<i>Variable</i>		<i>M</i>	<i>SD</i>	<i>F</i> _(59, 227)	<i>ICC</i> (1)	<i>ICC</i> (2)	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>
1.	Evaluation/learning	3.66	.39	2.11**	.24	.56	–				
2.	Discussing processes	3.05	.36	1.65**	.14	.38	.52**	–			
3.	Adaptation	3.64	.42	2.12**	.22	.51	.64**	.23	–		
4.	Feedback-seeking bh.	2.93	.52	2.14**	.24	.54	.58**	.49**	.35**	–	
5.	Reflector	3.99	.38	1.14	.05	.18	.21	.13	–.00	.00	–

Note: * *p* < .05; ** *p* < .01; two-tailed.

assessment showed a chi-square of 419.94 ($df = 208$), an adjusted goodness-of-fit index (AGFI) of .82, and a root mean square error of approximation (RMSEA) of .07 (Jöreskog & Sörbom, 1989).³ Modification index values indicated that we could improve the model by dropping four items, namely items 13, 14, 18, and 23, and theoretical examination supported their elimination. We dropped these items from the evaluation/learning scale, which now contains 14 items.⁴ After dropping these items, the same items loaded on the same factors across both studies, so the factor structure of Study 1 was replicated in Study 2.

We then compared the fit of the unidimensional model to the hypothesised two-factor structure (i.e. evaluation/learning and discussing processes as two separate but correlated constructs). In these models, the factors were allowed to correlate. For the unidimensional model, $\chi^2(135, N = 228) = 393.28$ ($p < .01$), AGFI = .77, RMSEA = .10; for the two-factor structure $\chi^2(134, N = 228) = 266.82$ ($p < .01$), AGFI = .84, RMSEA = .07. The significant improvement in fit of the two-factor solution over the unidimensional model ($\chi^2_{diff} = 126.46$, $df = 1$, $p < .01$) offers support for the convergent and discriminant validity of the scales, which was then further scrutinised (see next section).

Convergent and Discriminant Validity of the Scale Items. Following Netemeyer, Johnston, and Burton (1990), the scales used to operationalise the constructs were examined through the estimation of a measurement model. We used composite reliability (i.e. internal consistency, analogous to coefficient alpha) and variance extracted estimates (i.e. amount of variance captured by a construct in relation to the variance due to random measurement error) to assess the psychometric properties of the scaled measures. The composite reliabilities for the evaluation/learning, discussing processes, and adaptation scales were .90, .76, and .86, respectively (see Table 4; for a full description of how to compute composite reliabilities and variance extracted estimates, see Netemeyer et al., 1990). The t -values associated with the scale items ranged from 7.75 to 12.60 ($p < .01$), offering support for the convergent validity of the items in each scale.

³ We also compared the fit of the three-factor structure with the two- and one-factor structures. The three-factor model also had a significantly better fit than the one-factor model, but the improvement in model is virtually absent with respect to the two-factor model; no differences between the two models were found for the AGFI and RMSEA values. Hence, for reasons of parsimony, we chose the two-factor model over the other models. Calculations are available on request from the first author.

⁴ One item that was used in Study 1, namely item 23, was not used in Study 2, because initial analyses with the scale in Study 1, using varimax rotation, showed that this item should be left out. However, later analyses with oblimin rotation indicated that this item could have been retained.

TABLE 4
Measurement Properties (Study 2)

<i>Construct and indicators</i>	<i>Standardised loading</i>	<i>Reliability</i>	<i>Variance extracted estimate</i>
1. Evaluation/learning		.90 ^a	.39
λ_{x1}	.55	.30	
λ_{x3}	.57	.32	
λ_{x4}	.64	.41	
λ_{x5}	.51	.26	
λ_{x6}	.61	.37	
λ_{x7}	.72	.52	
λ_{x9}	.57	.32	
λ_{x10}	.65	.42	
λ_{x11}	.53	.28	
λ_{x12}	.52	.27	
λ_{x15}	.69	.48	
λ_{x16}	.72	.52	
λ_{x19}	.64	.41	
λ_{x20}	.72	.52	
2. Discussing processes		.76 ^a	.43
λ_{x24}	.54	.22	
λ_{x25}	.63	.40	
λ_{x26}	.76	.58	
λ_{x28}	.72	.52	
4. Adaptation		.86 ^a	.55
λ_{x1}	.76	.58	
λ_{x2}	.62	.38	
λ_{x3}	.81	.66	
λ_{x4}	.67	.45	
λ_{x5}	.82	.67	
5. Feedback-seeking bh.		.85 ^a	.54
λ_{x1}	.80	.64	
λ_{x2}	.71	.50	
λ_{x3}	.58	.34	
λ_{x4}	.71	.50	
λ_{x5}	.85	.72	
6. Reflector		.75 ^a	.51
λ_{x1}	.83	.69	
λ_{x2}	.53	.28	
λ_{x3}	.75	.56	

Note: ^a denotes composite reliabilities.

A test of discriminant validity (recommended by Fornell & Larcker, 1981, and described in full by Netemeyer et al., 1990) is to test whether the variance extracted estimates of the evaluation/learning and discussing processes scales exceed the square of the correlation between the two constructs. If this is the case, evidence of discriminant validity exists (Fornell & Larcker, 1981). The variance extracted estimates are .39 for evaluation/learning and

.43 for discussing processes. Both exceed the square of the correlations between the constructs ($\phi_{12} = .57$, $\phi_{12}^2 = .32$). This coefficient was also significantly less than 1, i.e. the confidence interval, plus or minus two standard errors, did not contain a value of 1 (SE of $\phi_{12} = .06$), which offers further support for the discriminant validity between the two constructs.

We then followed the same procedure regarding the other constructs measured in Study 2, namely adaptation, feedback-seeking behavior, and the reflector learning style. The composite reliability estimates were .86, .85, and .75, respectively. The variance extracted estimates (.55, .54, and .51) all exceeded the square of the correlations between the constructs (ϕ^2 's .12, .00, and .01, respectively).

Finally, in order to check whether evaluation/learning and discussing processes could be reliably distinguished from the other constructs used in Study 2, we compared the fit of the unidimensional model to the five-factor structure (i.e. evaluation/learning, discussing processes, adaptation, feedback-seeking behavior, and reflector) as five separate but correlated constructs. For the unidimensional model, $\chi^2(434, N = 225) = 1583.07$ ($p < .01$), AGFI = .59, RMSEA = .13; for the five-factor structure, $\chi^2(424, N = 225) = 801.93$ ($p < .01$), AGFI = .78, RMSEA = .06. The significant improvement in fit of the five-factor solution over the unidimensional model ($\chi^2_{\text{diff}} = 781.14$, $df = 10$, $p < .01$) supports the convergent and discriminant validity of the five scales, and also shows that the scales we chose to focus on in this study (evaluation/learning and discussing processes) can be discerned from other, related scales.

GENERAL DISCUSSION

Reflexivity can be seen as a key variable in team functioning (Schipper et al., 2003; Swift & West, 1998; West, 2000), yet research is scarce. The instrument to measure team reflexivity developed in the two studies presented here may help initiate further research. We started by writing items and included items describing different and broad situations in the reflexivity measure as we aimed to develop a questionnaire useful for different kinds of teams with a variety of tasks. We emphasised the reflection component and distinguished levels of reflection. We conducted a study to test the questionnaire in two samples of teams; the second was used as a confirmation sample. The initial evidence is sufficiently encouraging to suggest that reflexivity, and especially the two subscales of reflection, may be significant, measurable components of organisational and especially team behavior.

Factor analyses for the two different samples of Study 1 and Study 2 yielded similar results. Study 1 used a heterogeneous set of teams, whereas the teams in Study 2 were more homogeneous in terms of the team task. Two subscales of reflection were distinguished in both studies. Items in the

evaluation/learning scale emphasise evaluation of finished business and learning from previous actions and adaptations. These activities relate most to shallower levels of reflection. Discussing processes is aimed at a more “meta-level”, i.e. thinking about the way things are usually done in the team, reflecting on communication patterns within the team, and discussion of norms and values within the team. However, although theoretically three levels of reflection can be distinguished, we could not discern this third level of reflection from the others empirically. Both studies showed the two subscales to be internally consistent and reliable. Thus, despite a very different sample, the results of the study among school management teams replicated the factor structure and the high internal consistency of the reflection scales, and offered proof of discriminant validity. Further research is needed to assess whether the questionnaire is useful in yet other settings.

Although the components of reflexivity were highly correlated, they showed high internal consistency and were differentially related to some of the other variables. We showed convergent and discriminant validity for the scales in Study 2. Positive correlations with adaptation and feedback-seeking behavior were found, and as expected, lower correlations were found with the team level of proactive personality and the reflector learning style.

Limitations and Directions for Future Research

The results of this study represent an important first step in establishing and measuring reflexivity as a team-level construct. However, before firm conclusions can be drawn, additional conceptual and empirical work is needed to refine the measure, further tease out the relationships with related constructs, and to extend the implications.

The current study has several limitations. An important one is that respondents assessed the relevant variables themselves using questionnaires. Relationships between variables may therefore in part result from common method and common source variance. Future research might use multiple methods and sources to measure reflexivity in order to overcome this problem, such as coding reflexive behavior from videos of team interaction, interviews with team members and supervisors, and questionnaire data from peer teams, supervisors, or other external observers about the level of reflexivity in the focal team.

A related limitation concerns the cross-sectional design of both studies. This design did not allow us to assess predictive validity or establish causal relationships. Future research assessing the predictive validity of the reflexivity scale with a longitudinal design is needed. Also, this research might assess the test-retest reliability of the scale. In the current study, the expected third level of reflection could not be distinguished. We cannot rule out the possibility that this third level does not exist as a unified construct or that

it so rarely occurs that it will be hard to capture using questionnaires. Future research could perhaps develop new items intended to measure this level of reflection, or point out settings in which this kind of reflection would be most relevant.

A different problem may lie in possible self-selection of the teams. Teams with more interest in team communication and learning or those experiencing problems in that respect, might be more likely to participate. However, the sample displayed sufficient variation in both reflexivity and the other constructs measured. Also, all teams that were asked to participate, including those that ultimately did not, showed enthusiasm with respect to the subject. The reason for not participating was mostly lack of time due to pressing organisational matters. It is therefore not likely that biases were operative in this respect.

A final limitation is that situational contingency was difficult to assess as teams with routine tasks could not be included. Members of such teams often had a low educational level and supervisors of such teams indicated that the items in the reflexivity questionnaire would be too difficult as well as irrelevant for members of those teams. We therefore decided not to include teams with very routine tasks. The fact that the supervisors of such teams indicated that the items were not relevant with respect to the team task indicates that reflexivity may be most important to teams with moderate to highly non-routine tasks, as was already suggested by West et al. (1997). Unfortunately, no direct test of this proposition was possible. Future research might address for which tasks reflexivity is most important, as well as the ideal level of reflexivity for different kinds of tasks.

REFERENCES

- Allen, N.J. (1996). Affective reactions to the group and the organization. In M.A. West (Ed.), *Handbook of work group psychology* (pp. 371–396). Chichester: John Wiley & Sons.
- Ancona, D.G., & Caldwell, D.F. (1992). Bridging the boundary: External activity and performance in organizational teams. *Administrative Science Quarterly*, 37, 634–655.
- Argyris, C. (1992). *On organizational learning* (9th edn.). Malden, MA: Blackwell Publishers.
- Ashford, S.J., Blatt, R., & VandeWalle, D. (2003). Reflections on the looking glass: A review of research on feedback-seeking behavior in organizations. *Journal of Management*, 29, 773–799.
- Ashford, S.J., & Cummings, L.L. (1983). Feedback as an individual resource: Personal strategies of creating information. *Organizational Behaviour and Human Performance*, 32, 370–398.
- Bateman, T.S., & Crant, J.M. (1993). The proactive component of organizational behavior: A measure and correlates. *Journal of Organizational Behavior*, 14, 103–118.

- Bliese, P.D. (2000). Within-group agreement, non-independence and reliability: Implications for data aggregation and analysis. In K.J. Klein & W.J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations* (pp. 349–381). San Francisco, CA: Jossey-Bass.
- Carter, S.M., & West, M.A. (1998). Reflexivity, effectiveness and mental health in BBC production teams. *Small Group Research*, 29, 583–601.
- Cattell, R.B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1, 245–276.
- De Dreu, C. (2002). Team innovation and team effectiveness: The importance of minority dissent and reflexivity. *European Journal of Work and Organizational Psychology*, 3, 285–298.
- Den Hartog, D.N., Van Muijen, J.J., & Koopman, P.L. (1997). Transactional versus transformational leadership: An analysis of the MLQ. *Journal of Occupational and Organizational Psychology*, 70, 19–34.
- DeVellis, R.F. (1991). *Scale development: Theory and applications*. Newbury Park, CA: Sage.
- Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, 44, 350–383.
- Fornell, C., & Larcker, D.F. (1981). Evaluating structural equations models with observable variables and measurement error. *Journal of Marketing Research*, 18, 39–50.
- Frese, M., & Zapf, D. (1994). Action as the core of work psychology: A German approach. In H.C. Triandis, M.D. Dunette, & J.M. Hough (Eds.), *Handbook of industrial and organizational psychology* (Vol. 4, pp. 271–340). Palo Alto, CA: Consulting Psychologists Press.
- Hackman, J.R. (1987). The design of workteams. In J.W. Lorsch (Ed.), *Handbook of organizational behavior* (pp. 315–342). Englewood Cliffs, NJ: Prentice Hall.
- Honey, P., & Mumford, A. (1995). *The manual of learning styles* (3rd edn.). Maidenhead: Peter Honey.
- James, L.R. (1982). Aggregation bias in estimates of perceptual agreement. *Journal of Applied Psychology*, 67, 219–229.
- James, L.R., Demaree, R.G., & Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology*, 69, 85–98.
- Jöreskog, K.G., & Sörbom, D. (1989). *LISREL 7: A guide to the program and applications* (2nd edn.). Chicago, IL: SPSS Inc.
- Kenny, D.A., & La Voie, L. (1985). Separating individual and group effects. *Journal of Personality and Social Psychology*, 48, 339–348.
- Kolb, D.A. (1984). *Experiential learning: Experience as the source of learning and development* (1st edn.). Englewood Cliffs, NJ: Prentice Hall.
- Moreland, R.L., & Levine, J.M. (1992). Problem identification by groups. In S. Worchel, W. Wood, & J.A. Simpson (Eds.), *Group process and productivity* (pp. 17–47). Newbury Park, CA: Sage.
- Netemeyer, R.G., Johnston, M.W., & Burton, S. (1990). Analysis of role conflict and role ambiguity in a structural equations framework. *Journal of Applied Psychology*, 75, 148–157.
- Neuman, G.A., Wagner, S.H., & Christiansen, N.D. (1999). The relationship between work-team personality composition and the job performance of teams. *Group and Organization Management*, 24, 28–45.

- Nielson, R.P. (1993). Woolman's "I am we" triple-loop action-learning: Origin and application in organization ethics. *The Journal of Applied Behavioral Science*, 29, 117–138.
- Nunnally, J.C. (1976). *Psychometric theory*. New York: McGraw-Hill.
- Petzold, M. (1985). Kognitive stile: Definitionen, Klassifikationen und Relevanz eines psychologischen Konstrukts aus wissenschaftshistorischer Sicht. *Psychologie in Erziehung und Unterricht*, 32, 161–177.
- Schippers, M.C., Den Hartog, D.N., Koopman, P.L., & Wienk, J.A. (2003). Diversity and team outcomes: The moderating effects of outcome interdependence and group longevity, and the mediating effect of reflexivity. *Journal of Organizational Behavior*, 24, 779–802.
- Schwab, D.P. (1980). Construct validity in organizational behavior. In B.M. Staw & L.L. Cummings (Eds.), *Research in organizational behavior* (Vol. 2, pp. 3–43). Greenwich, CT: JAI Press.
- Senge, P.M. (1990). *The fifth discipline: The art and practice of the learning organization*. New York: Doubleday Currency.
- Snell, R., & Chak, A.M.K. (1998). The learning organization: Learning and empowerment for whom? *Management Learning*, 29, 337–364.
- Steiger, J.H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, 87, 245–251.
- Swift, T.A., & West, M.A. (1998). *Reflexivity and group processes: Research and practice*. Sheffield: The ESRC Centre for Organization and Innovation.
- Tjosvold, D. (1991). *Team organization: An enduring competitive advantage*. New York: John Wiley and Sons.
- Van Dyne, L., Graham, J.W., & Dienesch, R.M. (1994). Organizational citizenship behavior: Construct redefinition, measurement and validation. *Academy of Management Journal*, 37, 765–802.
- West, M.A. (1996). Reflexivity and work group effectiveness: A conceptual integration. In M.A. West (Ed.), *Handbook of work group psychology* (pp. 555–579). Chichester: John Wiley & Sons.
- West, M.A. (2000). Reflexivity, revolution and innovation in work teams. In M.M. Beyerlein, D.A. Johnson, & S.T. Beyerlein (Eds.), *Product development teams* (Vol. 5, pp. 1–29). Stamford, CT: JAI Press.
- West, M.A., Garrod, S., & Carletta, J. (1997). Group decision-making and effectiveness: Unexplored boundaries. In C.L. Cooper & S.E. Jackson (Eds.), *Creating tomorrow's organizations: A handbook for future research in organizational behavior* (pp. 293–316). Chichester: John Wiley & Sons.

APPENDIX: SURVEY SCALES

Reflection: Evaluation/learning

S1/S2

- 1/1. As a team we usually take well-considered decisions.
- 4/3. We review our methods of working as a result of changes in the environment.

- 5/4. We talk about different ways in which we can reach our objectives.
- 6/5. Problems are discussed only once they have become critical.^R
- 8/6. We examine the implications that changes in the environment may have for the aims of the team.
- 9/7. We work out what we can learn from past activities.
- 10/9. Before we get to work, we make sure everyone on the team has the same problem definition.
- 11/10. During task execution, we stop to assess whether the team is on the right track.
- 12/11. If a team member discovers a problem, he or she will talk about it with other team members.
- 13/12. We examine the long-term consequences of certain activities.
- 14/–. We question our objectives on a regular basis.
- 15/–. Problems are looked at from different points of view in this team.
- 18/15. We check whether our activities produced the expected results.
- 19/16. In this team the results of actions are evaluated.
- 20/–. We reflect on the question of whether a pattern can be discerned in events.
- 22/19. If things don't work out as planned, we consider what we can do about it.
- 23/–. If we are successful as a team, we take the time to analyse how we achieved this.
- 24/20. After certain activities are completed, we evaluate matters.
- 25/–. If things don't work out as they should, we take the time as a team to find the possible cause of the problems.

Reflection: Discussing Processes

S1/S2

- 26/24. The team often reviews its objectives.*
- 27/25. The methods used by the team to get the job done are often discussed.*
- 28/26. We regularly discuss whether the team is working effectively.*
- 32/28. The team often reviews whether it's getting the job done.*

Reflection: Discarded Items

S1/S2

- 2/2. We regularly reflect on the way in which we communicate.
- 7/–. Before we start on a task, we take time to discuss what the best working method is.
- /13. We question our objectives on a regular basis.
- /14. Problems are looked at from different points of view in this team.

- /18. We reflect on the question of whether a pattern can be discerned in events.
- 17/–. We regularly examine our contribution to the organisation.
- 21/–. This team is prepared to challenge the norms and values of the team.
- /23. If things don't work out as planned, we consider what we can do about it.
- 29/–. In this team we modify our objectives in the light of changing circumstances.*
- 34/–. The way decisions are made in this team is rarely altered.*
- 3/–. We regularly reflect on the way in which decisions are made.
- 30/–. Team strategies are rarely changed.*
- 31/27. We often discuss how well we communicate information.*
- 33/–. This team is prepared to challenge organisational practices and policies.*

Adaptation

1. After agreements have been made in this team, everyone does things a little differently.^R
2. In this team people keep to agreements.
3. In this team people have their own personal interpretation of agreements even when they are written down.^R
4. What we discuss corresponds with what we do subsequently.
5. After matters have been agreed, it turns out that different interpretations of the agreements exist among team members.^R

Feedback-seeking Behavior

1. We check on how satisfied others are with us.
2. We seek feedback on our methods.
3. We work out how well we are performing in comparison to other teams.
4. We ask for feedback from internal and external customers on our results.
5. We check how well we perform as a team.

Note: ^R Reversed items; S1 = items as described in Study 1; S2 = items as described in Study 2; – = not applicable; * Items drawn from Swift and West (1998).

Copyright of *Applied Psychology: An International Review* is the property of Blackwell Publishing Limited and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.